We claim:

1. A method for detecting misrepresentation of policy related information provided to an insurer by a policyholder where the information is used by the insurer in determining an amount of premium to be paid for insurance coverage provided to the policyholder, the method comprising:

selecting a plurality of insurance policies to process with a predictive model;

for each selected policy, deriving variables from policy related information provided by the policyholder in connection with the selected policy; and

for each selected policy, applying the derived variables of the policy to the predictive model to generate a model score indicating the relative likelihood of misrepresented information provided by the policyholder or an expected adjustment of the premium on the policy.

2. The method of claim 1, further comprising:

collecting training data including a plurality of insurance policies having misrepresented information and a plurality of policies having misrepresented information;

developing the predictive model from the training data; and

storing the predictive model.

3. The method of claim 1, further comprising:

converting the model score to a fraud score indicating a probability of fraud in the policy.

4. The method of claim 1, further comprising:

converting the model score to the expected adjustment of the premium on the policy.

5. The method of claim 1, wherein selecting a plurality of insurance policies further comprises:

for each policy, automatically determining start and end dates of a scoring period in which the determination of whether misrepresented policy information is to be determined.

6. The method of claim 1, further comprising determining the start and end dates of the scoring period which the policy has consistent and complete data.

7. The method of claim 6, further comprising:

responsive to a policy not having consistent or complete data in the scoring period, defining an exclusion code providing a reason that the policy was not selected.

8. The method of claim 6, wherein the insurance policies are workers' compensation insurance policies, and automatically determining start and end dates of the scoring period further comprises:

defining the start and end dates such that all audit adjustments are contained between the start and end dates.

9. The method of claim 1, wherein selecting a plurality of insurance policies further comprises:

for each policy, receiving a user defined scoring period to be scored for the policy; and

automatically selecting those policies having consistent and complete data in the respective user defined time period from which the variables for the predictive model may be derived.

10. The method of claim 9, further comprising:

responsive to a policy not having consistent or complete data in the user defined time period defining an exclusion code providing a reason that the policy was not selected.

11. The method of claim 9, further comprising:

responsive to a policy not having consistent or complete data in the user defined scoring period, automatically suggesting a scoring period in which the policy has consistent and complete data.

12. The method of claim 1, wherein deriving variables from policy related information further comprises:

determining a plurality of peer groups of which the selected policy is a member; and

for each peer group or set of peer groups of which the selected policy is a member, deriving variables from the policy information which attribute characteristics of the peer group or set of peer groups to the selected policy, or which compare the selected policy to other policies in the peer group or set of peer groups.

13. The method of claim 12, wherein the derived variables estimate the probability of a dichotomous outcome or a certain distributional statistic of a continuous quantity for a policy, based on the peer group(s) of which the policy is a member.

14. The method of claim 12, wherein deriving variables for the policy which compare the policy to other policies in its peer group(s) further comprises deriving variables that compare either at least one characteristic o of the policy with at least one corresponding characteristic of the policies in its peer group(s).

15. The method of claim 12, further comprising:

for each of the plurality of peer groups, storing in a lookup table group statistics for policy characteristics of the policies in the peer group; and

deriving the variables for a selected policy by determining the peer group to which the selected policy belongs and using the statistics for the policy characteristics for the peer group to derive the variables for the selected policy.

16. The method of claim 15, further comprising:

updating the lookup table for a peer group of the selected policy using policy information from the selected policy.

17. The method of claim 1, wherein deriving variables further comprises:

deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period.

18. The method of claim 17, wherein deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period further comprises:

deriving variables which quantify an amount or distribution of risk-related activities associated with the policy.

19. The method of claim 17, wherein deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period further comprises:

determining at least one measure which is a percentage change in a policy characteristic between the selected time period and the previous time period.

20. The method of claim 17, wherein deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period further comprises:

determining a vector of policy characteristics for the selected time period and a vector of the policy characteristics in the prior time period; and

determining a scalar measure of comparison between the two vectors.

21. The method of claim 20, wherein the scalar measure of comparison between the two vectors is computed as either a measure of distance between the two vectors or an angle measure between the two vectors.

22. The method of claim 17, wherein deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period further comprises:

determining a percent change in a payroll share in at least one employment classification in the selected time period relative to the previous time period.

23. The method of claim 17, wherein deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period further comprises:

determining a percent change in a payroll share in an exception group in the selected time period relative to the previous time period.

24. The method of claim 17, wherein deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period further comprises:

determining a vector distance between vectors of payroll percent shares in each of a plurality of employment classes in the selected time period and in the prior time period.

25. The method of claim 24, wherein the employment classes are SIC employment classes.

26. The method of claim 24, wherein the employment class groups are NCCI employment class groups.

27. The method of claim 24, wherein the employment class groups are rate-driven employment class groups.

28. The method of claim 24, wherein the employment class groups are data-driven employment class groups, each group including employment classes that are likely to appear together in payroll reports.

29. The method of claim 17, wherein deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period further comprises:

determining a percent change in a number of claims filed on the policy in the selected time period relative to number of claims filed on the policy in the prior time period.

30. The method of claim 17, wherein deriving variables from the policy information which compare the selected policy in a selected time period with the selected policy in a time period prior to the selected time period further comprises:

determining a vector distance between a first vector of the number of claims filed in the selected time period for each of a plurality of injury types and a second vector of the number of claims filed in the prior time period in each of the plurality of injury types.

31. The method of claim 1, wherein the insurance policies are workers' compensation insurance policies and the policy relative information from which the variables for assessing the policies are derived includes payroll reports for the policyholder.

32. The method of claim 1, further comprising:

deriving direct policy variables which measure characteristics of the policyholder or the policy itself without comparison to other policies or the same policy in a prior time period.

33. The method of claim 32 wherein the direct policy variables are selected from the group consisting of:

type of company of the policyholder;

location of the policyholder;

number of employees of the policyholder;

number of policy cancellations;

age of the policy;

industry type of the policyholder;

amount of payroll reported by the policyholder; and

distribution of payroll reported by the policyholder with respect to at least one employment class.

34. The method of claim 1, further comprising:

deriving direct claim variables which measure characteristics of claims filed on policy.

35. The method of claim 34 wherein the direct claim variables are selected from the group consisting of:

number of claims filed during the selected time period;

dollar amount of claims filed during the selected time period;

type of claims filed during the selected time period;

number of claims filed during the selected time period relative to amount of premium paid during the selected time period; and

number of claims filed during the selected time period relative to a size of payroll during the selected time period.

36. The method of claim 1, further comprising deriving variables that measure the probability of fraud in the policy conditionally based on at least one policy characteristic of the policy.

37. The method of claim 1, further comprising:

applying the policy to a plurality of decision rules which identify specific inconsistent or suspicious policy facts related to the policy, to generate an output indicating which decision rules were violated by the policy.

38. The method of claim 37, wherein the decision rules are derived from statistical analysis of insurance policies of at least one insurer which have been determined to contain misrepresented policy information.

39. The method of claim 37, wherein the insurance policies are workers' compensation insurance policies and wherein the decision rules are selected from a group consisting of:

a decision rule that identifies as potentially fraudulent a policy that has an employment class code on a claim with an injury date during the selected time period but the employment class code for the claim is not included in payroll reports for the policy during the selected time period;

a decision rule that identifies as potentially fraudulent a policy that reports zero payroll during the selected time period but for which one or more certificates of insurance were issued during the selected time period;

a decision rule that identifies as potentially fraudulent a policy that reports zero payroll during the selected time period but which has at least one claim with an injury date during the selected time period;

a decision rule that identifies as potentially fraudulent a policy with an officer who is currently or was selectedly an officer on a different policy and where the new policy has a lower experience modification factor than the prior policy; and

a decision rule that identifies as potentially fraudulent a policy that has an employment class code on a claim and for which no premium was reported at the time the claim was opened

40. The method of claim [illegible], further comprising:

for each selected policy, determining at least one variable which significantly contributes to the model score for the policy; and

outputting a reason for the model score with respect to the determined at least one variable.

41. The method of claim 40, wherein the insurance policies are workers' compensation insurance policies, and wherein the significant variable is selected from a group consisting of:

an indication of whether the policy has been previously audited;

an indication of whether a reported payroll has been adjusted;

a number of employment class codes in at least one payroll report of the policyholder during the selected time interval;

a type of company of the policyholder;

an age of the policy;

a size of payroll of the policyholder;

a size of a premium paid on the policy;

an industry classification code of the policyholder;

a distribution of payroll in at least one payroll report of the policyholder during the selected time interval;

a percent payroll share in a low rated employment class code;

a change in a distribution of payroll in at least one payroll report of the policyholder during the selected time interval compared with the prior time period;

a change in an exception group payroll share in at least one payroll report of the policyholder during the selected time interval compared with the prior time period;

a payroll share in a group of agriculture related employment classes;

a payroll share in a group of construction related employment classes;

a payroll share in a group of manufacturing related employment classes;

a payroll share in a group of government related employment classes;

a payroll share in at least one clerical employment classes;

a number of prior cancellations of the policy;

a ratio of the number of claims made on the policy to a size of the payroll of the policyholder; and

a number of claims on the policy during the selected time interval.

42. A method for training a neural network on a plurality of observations to score the observations on a dependent variable, each observation including an independent variable having an original value that is highly correlated with the dependent variable, so as to calibrate the influence of the independent variable on scores, the method comprising:

for each of the plurality of observations, setting the independent variable to a randomly selected value, and providing the observations to the neural network a first time, wherein the neural network establishes connection weights based on the provided observations to output an un-calibrated score for an observation; and

for each of the plurality of observations, setting the independent variable to its original value in the observation, and providing the observations to the neural network a second time, wherein the neural network adjusts the connection weights to calibrate the output scores with respect to the independent variable.

43. The method of claim 42, wherein the independent variable is a Boolean variable having two defined values, and the randomly set value is between the two defines values of the Boolean variable.

44. The method of claim 42, wherein the independent variable is a continuous variable having a range of values, and the randomly set value is within the range of values.

45. A method of estimating a quantity corresponding to a set of entities grouped using one or more hierarchical categories, the method comprising:

determining an estimate of the quantity for a first category corresponding to the highest level of the hierarchy; and

for each subsequent category representing a current, lower level of the hierarchy, adjusting the estimate of the quantity using an estimate for the current level and the estimate of the higher level.

46. The method of claim 45, wherein the quantity being estimated is a risk factor, and each category of the hierarchy has a value for the risk factor.

47. The method of claim 45, wherein the hierarchy of categories are Standard Industry Classification codes (SIC), and the quantity being estimated is risk factor associated with each SIC code.

48. The method of claim 45, wherein adjusting the estimate of the quantity comprises applying a Bayesian adjustment to the estimate using the estimate for the current level of the hierarchy and the estimate of the quantity from the higher level.

49. A system for detecting premium fraud in an insurance policy, comprising:

a database of insurance policies, each policy associated with a policyholder and having policy related data;

a policy selection process that selects from the database a number of policies for scoring;

a variable derivation process that derives for each of the selected policies variables associated with the policyholder of the policy for comparing the policy to peer group policies, and variables for comparing the policy in a selected time period with the policy a time period prior to the selected time period; and

a fraud detection module that receives for each policy the derived variables and generates a score indicating the likelihood of misrepresentation of policy information by the policyholder of the policy.

50. The method of claim 49, wherein the fraud detection module further comprises:

a predictive model that generates a model score indicating a relative likelihood of misrepresentation of policy information by the policyholder; and

a post scoring process that converts the model score into the fraud score indicating a probability of misrepresentation of policy information.

51. The system of claim 50, wherein the post scoring process converts the model score into an expected adjustment of premium for a policy.

52. The system of claim 50, further comprising:

a rule-based process that applies a plurality of rules to a selected policy to identify policies suspected of premium fraud based on inconsistent or incomplete policy related information.

53. A method for determining a usage strategy for processing insurance policies suspected of premium fraud, the suspected policies selected from a plurality of insurance policies, the method comprising:

establishing a frequency for scoring the plurality of insurance policies to obtain for each policy a score indicating a relative likelihood of premium fraud in the policy;

establishing a ranking function for ranking the scored policies; and

establishing a plurality of threshold scores, and for each threshold score, defining an audit action for performing on policies which have a score exceeding the threshold score, but not exceeding a next greater threshold score.

54. The method of claim 53, wherein establishing a ranking function for ranking the scored policies further comprises:

ranking the scored policies according to their scores.

55. The method of claim 53, wherein establishing a ranking function for ranking the scored policies further comprises:

ranking the scored policies according to an expected adjusted premium.

56. The method of claim 53, wherein establishing a plurality of threshold scores further comprises:

establishing a first threshold score for selecting for a desk audit those policies having a score exceeding the first threshold score; and

establishing a second threshold score for selecting for a field audit those policies having a score exceeding the second threshold score, wherein the second threshold score is greater than the first threshold score.

57. The method of claim 53, further comprising:

establishing a set of rules for identifying policies suspected of premium fraud.

58. The method of claim 53, further comprising:

establishing a plurality of reason codes, each reason code providing an explanation for a policy receiving a score; and

establishing for each of number of reason codes, at least one audit action to be taken in response to a policy having a score which produces the reason code.

59. A method for processing insurance policies suspected of premium fraud, the method comprising:

scoring each of a plurality of insurance policies with predictive model to generate for each policy a score indicating a relative likelihood of premium fraud;

ranking the scored policies according to the scores;

selecting for a desk audit those policies having a score exceeding a first threshold score; and

selecting for a field audit those policies having a score exceeding a second threshold score, wherein the second threshold score is greater than the first threshold score.

60. A method for processing insurance policies suspected of premium fraud, the method comprising:

scoring each of a plurality of insurance policies with predictive model to generate for each policy a score indicating a relative likelihood of premium fraud;

determining for each scored policy an expected premium adjustment;

ranking the scored policies according to their expected premium adjustments;

selecting for a desk audit those policies having an expected premium adjustment exceeding a first threshold amount; and

selecting for a field audit those policies having a expected premium adjustment exceeding a second threshold amount, wherein the second threshold amount is greater than the first threshold amount.

61. A method of developing a predictive model of insurance premium fraud, the method comprising:

collecting from at least one insurance company policy information for a plurality of insurance policies;

determining for each policy a scoring period for scoring the policy;

selecting a training set of policies;

deriving for each policy in the training set a plurality of variables from the policy information and from other information relevant to policy premiums;

applying the derived variables to an untrained predictive model to train the predictive model to produce a measure with respect to whether the policies are fraudulent or non-fraudulent during their respective scoring periods ; and

selecting a subset of the derived variables for the using in the predictive model, which variables significantly contribute to a prediction of whether a policy is fraudulent during its scoring period.

62. The method of claim 61, wherein the insurance policies are workers' compensation insurance policies, further comprising:

excluding from the training set policies for which no payroll is reported during the scoring period for the policy.

63. The method of claim 61, further comprising:

tagging each of the policies to indicate whether the policy is fraudulent, non-fraudulent, or indeterminate; and

excluding from the training set policies which are tagged as indeterminate.

64. The method of claim 61, further comprising:

for each policy in the training set, providing a random value for the previously audited variable, and applying the derived variables and the random value of the previously audited variable to the predictive model; and

for each policy in the training set, providing an actual value for the previously audited variable indicating whether the policy was previously audited for the scoring period, and applying the derived variables and the actual value of the previously audited variable to calibrate the scores produced by the predictive model.